

Universal Multiple-Octet Coded Character Set
 International Organization for Standardization
 Organisation internationale de normalisation
 Международная организация по стандартизации

L2/03-133R

Doc Type: Working Group Document
Title: Revised proposal to encode the ARABIC ZWARAKAY in the UCS
Source: Michael Everson and Roozbeh Pournader
Status: Individual Contribution
Replaces: N2581R
Action: For consideration by JTC1/SC2/WG2 and UTC
Date: 2003-05-29

In December 2002 the United Nations Development Programme Country Office for Afghanistan commissioned a report on the language support required for Pashto and Dari, Afghanistan's official languages. As a part of this report, we undertook a review of the character requirements of these languages. One character used in Pashto educational and lexicographical texts, is missing from the UCS. We propose its addition here, at U+0659.

a diacritical mark *zi* (zawar) زور
 representing (a)
 see (شور) شور (zawar) زور
 a diacritical mark *zi* (zwarakay) زورکی
 mark representing (ə).

On the left, the entries for *zwar* [a] for *zwarakay* [ə] from Rahimi and Rohi's *Pashto-English Dictionary*, 1979, p. 246. On the right, the entry for *zwarakay* [ə] from *Pashto-Pashto Descriptive Dictionary*, 4 volumes (Department of Linguistics, Institute of Languages and Literature, Academy of Sciences of Afghanistan, 1979-1987), vol. 3, p. 2401.

زورکی (zwarakay) زورم .
 خفیف زور ، سپک زور ؛ په
 پښتو او ځینو نورو ژبو کې یو
 حرکت دی چې نه زور (فتحه) دی
 اونه زپر (کسره) ؛ لکه دلپسه
 [psə, pusə] په کلمه کې د (س) حرکت
 اود انگلیسي په *ago* کې لومړی حرکت
 اود *China* په کلمه کې دویم حرکت .
 د دې حرکت د تمثیل دپاره معمولاً په فونیمیک
 او فونیتیک سیستم کې د [ə] سمبول
 کارېږي .

The vol. 4 text to the right reads "m. sg. Weak *zwar*, light *zwar*; in Pashto and some other languages, is a diacritical mark that is neither a *zwar* (*fatha*) nor a *zer* (*kasra*); like the sound of (س) in the word پسه [psə, pusə] and the first sound in *ago* and the last sound in the word *China* in English. Usually, the [ə] symbol is used to represent this sound in phonemic and phonetic systems.

نورې نڅنې : ۸ د ساکن توري د ښوولو دپاره او — د زورکي دپاره استعمال شوي دي .
 په موسيقي کې (و) دخالي نڅنې ، (و) دسم نڅنې او (و) دتالي نڅنې ده . همدارنگه د لگه ،
 گر ، پلټ ، درت ، برام ، لگه برام اوله برام ، درت برام او اُنډرت د ښوولو دپاره په
 ترتيب سر دغه نڅنې کار شوي دي : ا ، ٤ ، ٥ ، آ ، ٥ ، ٥

Front matter from vol. 3, p. م (above) and vol. 4, p. غ (right) from *Pashto-Pashto Descriptive Dictionary*, 4 volumes (Department of Linguistics, Institute of Languages and Literature, Academy of Sciences of Afghanistan, 1979-1987). The vol. 3 text reads "Other Symbols: SUKUN is used to show the *sukun* letter and ZWARAKAY is used for *zwarakay*."

نورې نڅنې : په موسيقي کې (و) دخالي نڅنې ، (و) دسم نڅنې او (و) دتالي نڅنې ده . همدارنگه د لگه ،
 گر ، پلټ ، درت ، برام ، لگه برام اوله برام ، درت برام او اُنډرت نڅنې دي : ا ، ٤ ، ٥ ، آ ، ٥ ، ٥
 ۸ يا ۹ د ساکن توري نڅنې
 — زورکی
 ٤ د (و) يعني ايضاً پځای
 ٥ د ستونيز تمېدنگ (glottal stop) نڅنې ؛ لکه په *gur'an* کې .
 * د شرې په پای کې پايېک وگورئ .
 ... اوداسې نور ، دوام لري ، صرف نظر
 # ډيز
 b بيول
 ٤ پکار ، نانوريل

زورکی

(یا حرکت مخصوصه)

در افغانی علاوه بر حرکات ثلاثه یک حرکت چهارم نیز میباشد که از خواص افغانی شمرده میشود، مثلاً شل (بیت) شل (صد) دی (هست) اگر به حرکت مخصوصه یعنی (زورکی)

ادا نشود معنی تغییر پیدا میکند مثلاً: دی (به فتح دال) بمعنی آن یکرد و به زورکی معنی، هست را میدهد. چون دی دی (آن یکرد هست) و برای این حرکت چنین علامه (-) مقرر نمودیم که زبر افقی میباشد، زبر یعنی فتحه قدره خمیدگی دارد و این راست میباشد زبر (-) زورکی (-) رعایت این حرکت خیلی ضروریست ورنه معنی با لکل غلط میشود

Examples of ZWARAKAY in use: بَل

[bəl] ‘other’ vs بَل [bal] ‘ablaze’.

زِرَه [zəɾə] ‘old f.’ vs زِرَه [zɾə] ‘heart’.

Text from Majles-e ta’lif-e pashto (The Pashto Compilation Assembly), *Mo’allem-e pashto* (‘The Pashto Teacher’) Kandahar 1312 (1933), showing minimal pairs with fatha and zwarakay:

In Afghani [Pashto] there is a fourth vowel in addition to the main three vowels, which is specific to Afghani, for example شل (twenty), سَل (hundred), and دِی (is), if not pronounced using the special vowel that is *zwarakay*, will have different meanings. For example: دی (with *fatha* over *dal*) means ‘that one man’, while with *zwarakay*, it will mean ‘is’. Like دِی دِی (That one man is [exists]). And for this vowel, we prescribe the a sign like “-”, that is a horizontal *zebar*. *Zebar* or *fatha* has some curvature [sic], but this is straight. *Zebar*: “-”, *zwarakay*: “-”. Observing this vowel is highly necessary, or otherwise the meaning will become wrong altogether.

0659



ARABIC ZWARAKAY

A. Administrative

1. Title

Revised proposal to encode the ARABIC ZWARAKAY in the UCS.

2. Requester’s name

Michael Everson and Roozbeh Pournader

3. Requester type (Member body/Liaison/Individual contribution)

Individual contribution.

4. Submission date

2003-05-29

5. Requester’s reference (if applicable)

N2581R

6. Choose one of the following:

6a. This is a complete proposal

Yes.

6b. More information will be provided later

No.

B. Technical -- General

1. Choose one of the following:

1a. This proposal is for a new script (set of characters)

No.

Proposed name of script

1b. The proposal is for addition of character(s) to an existing block

Yes.

1b. Name of the existing block

Arabic.

2. Number of characters in proposal

1

3. Proposed category (see section II, Character Categories)

Category A.

4a. Proposed Level of Implementation (1, 2 or 3) (see clause 14, ISO/IEC 10646-1: 2000)

Level 3.

4b. Is a rationale provided for the choice?

Yes.

4c. If YES, reference

Combining characters.

5a. Is a repertoire including character names provided?

Yes.

5b. If YES, are the names in accordance with the character naming guidelines in Annex L of ISO/IEC 10646-1: 2000?

Yes.

5c. Are the character shapes attached in a legible form suitable for review?

Yes.

6a. Who will provide the appropriate computerized font (ordered preference: True Type, or PostScript format) for publishing the standard?

Michael Everson.

6b. If available now, identify source(s) for the font (include address, e-mail, ftp-site, etc.) and indicate the tools used:

Michael Everson, Fontographer.

7a. Are references (to other character sets, dictionaries, descriptive texts etc.) provided?

Yes.

7b. Are published examples of use (such as samples from newspapers, magazines, or other sources) of proposed characters attached?

No.

8. Does the proposal address other aspects of character data processing (if applicable) such as input, presentation, sorting, searching, indexing, transliteration etc. (if yes please enclose information)?

No.

9. Submitters are invited to provide any additional information about Properties of the proposed Character(s) or Script that will assist in correct understanding of and correct linguistic processing of the proposed character(s) or script. Examples of such properties are: Casing information, Numeric information, Currency information, Display behaviour information such as line breaks, widths etc., Combining behaviour, Spacing behaviour, Directional behaviour, Default Collation behaviour, relevance in Mark Up contexts, Compatibility equivalence and other Unicode normalization related information. See the Unicode standard at <http://www.unicode.org> for such information on other scripts. Also see Unicode Character Database <http://www.unicode.org/Public/UNIDATA/UnicodeCharacterDatabase.html> and associated Unicode Technical Reports for information needed for consideration by the Unicode Technical Committee for inclusion in the Unicode Standard.

This character is very similar in behaviour to the ARABIC FATHA (U+064E), and may possibly appear together with an ARABIC SHADDA over a base letter. We recommend the following Unicode Character Properties to be assigned to the proposed character:

General Category: Mn

Canonical Combining Class: 37

Bidi Class: NSM

Bidi Mirrored: N

C. Technical -- Justification

1. Has this proposal for addition of character(s) been submitted before? If YES, explain.

No.

2a. Has contact been made to members of the user community (for example: National Body, user groups of the script or characters, other experts, etc.)?

Yes.

2b. If YES, with whom?

The Pashto experts Dr Habibullah Rafi (the editor of the popular magazine *Kelid* published in Kabul), and Mr Said Marjan Zazai (a Pashto computerization expert from the Afghan Assistance Coordination Authority)

2c. If YES, available relevant documents

3. Information on the user community for the proposed characters (for example: size, demographics, information technology use, or publishing use) is included?

Pashto-speaking educators among others.

4a. The context of use for the proposed characters (type of use; common or rare)

To represent the schwa sound in Pashto.

4b. Reference

5a. Are the proposed characters in current use by the user community?

Yes.

5b. If YES, where?

Educational materials.

6a. After giving due considerations to the principles in Principles and Procedures document (a WG 2 standing document) must the proposed characters be entirely in the BMP?

Yes.

6b. If YES, is a rationale provided?

Yes.

6c. If YES, reference

Keep with other Arabic signs.

7. Should the proposed characters be kept together in a contiguous range (rather than being scattered)?

8a. Can any of the proposed characters be considered a presentation form of an existing character or character sequence?

No.

8b. If YES, is a rationale for its inclusion provided?

8c. If YES, reference

9a. Can any of the proposed characters be encoded using a composed character sequence of either existing characters or other proposed characters?

No.

9b. If YES, is a rationale for its inclusion provided?

9c. If YES, reference

10a. Can any of the proposed character(s) be considered to be similar (in appearance or function) to an existing character?

Yes. There are other more-or-less horizontal bars in the standard.

10b. If YES, is a rationale for its inclusion provided?

Yes.

10c. If YES, reference

This character behaves as other Arabic combining vowel signs do.

11a. Does the proposal include use of combining characters and/or use of composite sequences (see clauses 4.12 and 4.14 in ISO/IEC 10646-1: 2000)?

Yes.

11b. If YES, is a rationale for such use provided?

Yes.

11c. If YES, reference

It is a combining character.

12a. Is a list of composite sequences and their corresponding glyph images (graphic symbols) provided?

No.

12b. If YES, reference

13a. Does the proposal contain characters with any special properties such as control function or similar semantics?

No.

13b. If YES, describe in detail (include attachment if necessary)

14a. Does the proposal contain any Ideographic compatibility character(s)?

No.

14b. If YES, is the equivalent corresponding unified ideographic character(s) identified?